Development of Hybrid Decision Support Framework for IoT Based Big Data Environment

A
SYNOPSIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

COMPUTER SCIENCE

BY
RAJESH SHARMA

UNDER THE SUPERVISION OF

PROF. PREETVANTI SINGH

DEPARTMENT OF PHYSICS AND COMPUTER SCIENCE

FACULTY OF SCIENCE

DAYALBAGH EDUCATIONAL INSTITUTE (DEEMED UNIVERSITY)

DAYALBAGH, AGRA-282005

2022

INTRODUCTION

In today's digital world, huge amount of data is being generated at high speeds by a variety of sources (like social media and machine logs), new technologies and paradigms (such as Internet of Things) and smart digital devices. This massive amount of data, growing exponentially at an unprecedented rate is referred to as Big Data. It is a blanket term for the non-traditional strategies and technologies required to organize, process, and gather insights from large datasets. Benefits of big data include (Oguntimilehin & Ademola, 2014):

- *Easy utilization of outside intelligence:* While taking business decisions to improve its operations, access to social data enables organizations to fine tune their business strategies.
- *Sophisticated analysis of customers' trends:* Big data allows companies to match customer expectation.
- *Identification of potential risks:* Big data can help the organizations develop new risk management solutions and get smarter strategies.
- *Real time analysis:* Big data technologies allow getting actionable data that can be used to analyze the customer one-on-one in real time.
- *Development of dynamic websites:* Big data analysis helps in personalizing the look, content and feels of the websites that suits every visitor.

The main attributes of Big Data are:

a. **Volume (The quantity of data):** The size of big data is usually much more than terabytes or even Petabytes because the data is created from different sources.

b. **Velocity (The rate of data generation and transmission):** Velocity in big data deals with the speed of the data coming from different sources and also with the speed at which the data moves.

c. **Variety (Structured, semi-structured and unstructured data):** Variety refers to the types of data collected via sensors, smartphones, or social networks. Such data types include image, text, audio, video, data logs and others in structured, semi-structured or unstructured format.

d. **Value (Important results from the filtered data):** Value helps in discovering vast hidden values from huge datasets with various types and rapid generation. User can run queries against the data

stored to deduct important results. These reports help to discover the business movements according to which decision makers can change their strategies.

e. **Veracity (The trust and integrity):** Veracity refers to the biases, noise, abnormality in data, uncertainty due to data inconsistency and incompetence, ambiguities, latency, deception and model approximations for the data being stored and analyzed.

These V's of big data are crucial for gathering, storing, analyzing, managing and consuming huge sets of information. Because of the size and complexity, these sets go beyond the traditional data computation methods.

The big data can be better understood based on the following classification:

A. **Data sources:** Big data comes from three primary sources:

1. **Web and Social Media:** Social media platforms have nowadays become the most popular source of data, where the data comes from the likes, tweets, comments, video uploads, and the media that are uploaded and shared. This data provides valuable insights into consumer sentiment and behaviour which can be useful in marketing analytics. Internet or the public web is also the most widely used and best-known source of big data that is easily accessible. This is used by organizations to have integrated and detailed customer-level data for enterprise analytics.
2. **Machine-generated:** Operations of various industrial and medical equipment, network devices (including computers, mobile devices, embedded systems), Internet of Things, and web logs (which track user behavior) are the common sources of generating machine data, usually without human intervention. Combining machine data with other enterprise data for analysis provides new views and insights for business operations and activities.
3. **Transactional data:** Transaction data is the data directly derived as a result of transactions. It contains a time dimension and thus becomes less relevant over the time. This is generated from all the daily transactions that take place both online and offline. Transactional data describes an internal or external event which takes place as the organization conducts business. It supports ongoing business operations and can be financial, logistical or any business-related processes. Rather than being the object of transactions like the identity of the customer, it is a reference data describing the time, place, prices, payment methods, discount values, and quantities related to that particular

transaction, usually at the point of sale. It is grouped with its associated and references master data such as product information and billing sources.

B. Content format: Big data can be of following types:

- **Structured data:** This type of data has a particular format. It can be managed in a data warehouse where data is grouped into a relational scheme. Queries can be generated to reach at useful information based on an organization's parameters and operational needs using SQL.
- **Semi-structured data:** This type of data does not conform with the formal structure of data models associated with relational databases or other forms of data tables but contain tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Examples include emails, or NoSQL.
- **Unstructured data:** Unstructured data is the data that is not arranged according to a pre-set data model or is not organized in a pre-defined manner. This type of data is random and difficult to analyze. Examples include images, audio and video files, text messages, location information, videos, and social media data.

C. Data Stores: Big data can have different data stores:

1. **Graph Based:** A graph database, such as Neo4j, is designed to store and represent data that use a graph model with edges, nodes, and properties related to one another through relations. These help in uncovering patterns and new insights by understanding the connections between data and the meaning of these links. For example, the graph based dataset can be useful in making link analysis of Twitter, Facebook and LinkedIn data.
2. **Column-Oriented:** A column-oriented database, such as HBase, stores its content in columns aside from rows, with attribute values belonging to the same column stored contiguously. Column-oriented is different from classical database systems that store entire rows one after the other. The columnar database is preferred for analytical applications that requires real-time information exchange or have frequently changing content serving.
3. **Document-Oriented:** A document-oriented database is a database that deals with documents. It plays an important role in aggregating data from documents and getting them into a searchable and organized form. It supports difficult data forms in several standard formats, such as JSON, XML,

and binary forms (such as MS Word and PDF). A document-oriented data store (such as CouchDB and MongoDB) is same as a record or row in a relational database however is more flexible and can retrieve documents based on their contents.

4. **Key-Value:** Key-value database is a non-relational database system (for example Riak and Oracle NoSQL) that stores and accesses data designed to scale to a very large size. The database stores data as a collection of key-value pairs, where a key serves as a unique identifier and clients may input queried values according to the keys. Both keys and values can be anything ranging from simple objects to complex compound objects. These data sets are created from diverse sources such as sensors that gather weather information, magazines and newspaper articles, web logs, military records and so on.

D. Big Data Pre-Processing: Before the data can be used for analysis, it requires pre-processing that includes:

- **Noise Identification:** The veracity of data depends on identifying and removing meaningless information present in data. The noise can also be random fluctuations in data values that harm analytic predictions.
- **Data Cleaning:** It is the process of identifying incomplete and unreasonable data that includes detecting, correcting, or removing corrupt/inaccurate records in raw data by removing typographical errors and validating values against a known list of entities. Noisy data is also smoothened out or removed in this stage.
- **Data Normalization:** It is the process of transforming data into a form suitable for analysis. For example, skewness in data is removed by transforming all variables in data to specific range for better analytics processing.
- **Data Transformation:** This is the method of structuring database schema to minimize redundancy. It involves converting data from one format or structure to another format suitable for integration, warehousing and wrangling.
- **Data Integration:** This involves combining data from multiple sources into an integrated view. As volume of data increases algorithms are applied, data integration becomes more critical.

E. Data Processing: Big data can be processed in following ways:

1. **Real-Time Processing:** Real-time processing means that data is processed as it arrives and either a business user gets consumable insights without exceeding a time period allocated for decision-making, or an analytical system triggers an action or a notification.

   Real-time processing paradigm deals with velocity of Big Data such as processing streaming data with low latency. Real-time processing can be thought of an infinite sequence of small batch processing where the information is in memory instead of secondary storages. Processed data is written to optimize analytical data store for analytics, business intelligence, and real-time dashboard visualization. This type of processing can help in keeping up-to-date data on what is happening *right now*, for example, defining current or trending topics at Twitter. Other examples of real-time processing include managing transportation, energy supply, and garbage collection in smart cities; or using data gathered from emergency management sources, citizens' usage of social networks, and mobile devices for disaster management; or using factories' sensors for quality control and product optimization. Apache Flink, Apache Spark and Apache Storm are some real-time distributed tools for processing data streams.
2. **Batch processing:** Batch processing paradigm deals with volume of big data. The system allows for the scaling of applications across large clusters of machines comprising thousands of nodes. It is the solution to process large volume of static data. This paradigm does not take into account new data once a batch processing has started. The main feature of batch processing system is the scalability. To achieve high scalability and deal with the volume problem, batch processing uses a parallel distributed processing framework such as MapReduce. It has several advantages such as allowing for a simple and unifying view of data, is inherently scalable and hides the complexity of programming distributed software, which is challenging because of potential hardware failures, fluctuations in network quality, and device heterogeneity.
3. **Hybrid Processing:** Many application domains require the combination of batch and real-time processing paradigms. This is achieved through a hybrid model, known as Lambda Architecture. It contains the following:

- **Batch layer (batch processing):** manages the master dataset, which is not changeable and is stored in a distributed file system;

- **Serving layer (batch results):** loads and exposes the batch views in a data store for querying;
- **Speed layer (real-time processing):** deals only with new data that require low latency. To obtain a complete result, the batch and real-time views must be queried, and the results be merged together.
- **Combination layer:** Synchronization, results composition, and other non-trivial issues have to be addressed at this stage.

Chen & Fu (2015), Harb, Mroue, Mansour, Nasser & Motta Cruz (2020), Verma, Malhotra & Singh (2020) and Renjith, Sreekumar & Jathavedan (2021) developed various algorithms to analyze the mountain of the data efficiently.

IoT generated Big Data

IoT, a world-wide network of interconnected objects, is pointing towards productive and new directions for a whole generation of information devices. Nowadays, tremendous volumes of data are generated by IoT devices which demands new architectures and technologies to capture and process this big data for enhanced insights and decision making.

There are various benefits of IoT based big data:

- **Real-time monitoring:** Big data gathered by connected devices can be used in real-time operations, for example, tracking physical activities (say counting steps) monitoring heart rate, measuring blood pressure, or sugar, controlling production machinery, or monitoring cattle and plants.
- **Data analysis:** IoT based big data can be used to identify trends and tendencies, reveal unseen patterns and find hidden information and correlations.
- **Process control and optimization**: Data that comes from sensors gives additional context to reveal non-trivial issues affecting performance and optimize processes, for example, tracking traffic load in various dates and times to work out the recommendations aimed at traffic optimization.
- **Predictive maintenance:** IoT based big data can be a reliable source to predict risks, proactively identify potentially dangerous conditions, for example, monitoring patients' state and identifying risks of say diabetes, or heart attacks to take timely measures, or predicting equipment failures.

Thus, IoT generated data can be beneficial in getting by with quick analysis, or by performing deeper data processing.

The steps of IoT generated big data processing are as follows (Figure 1):

1. A large amount of unstructured data is generated by IoT devices which are collected in the *Data Lakes*. This data largely depends on heavily volume, variety and velocity and is stored in its natural format. The data when cleaned and structured, is stored in a *Data Warehouse*.
2. In the data warehouse, which is a shared distributed database, the huge amount of data is stored in big data files.
3. This stored data is analyzed using analytic tools.
4. Finally, the results (reports or queries) are generated.

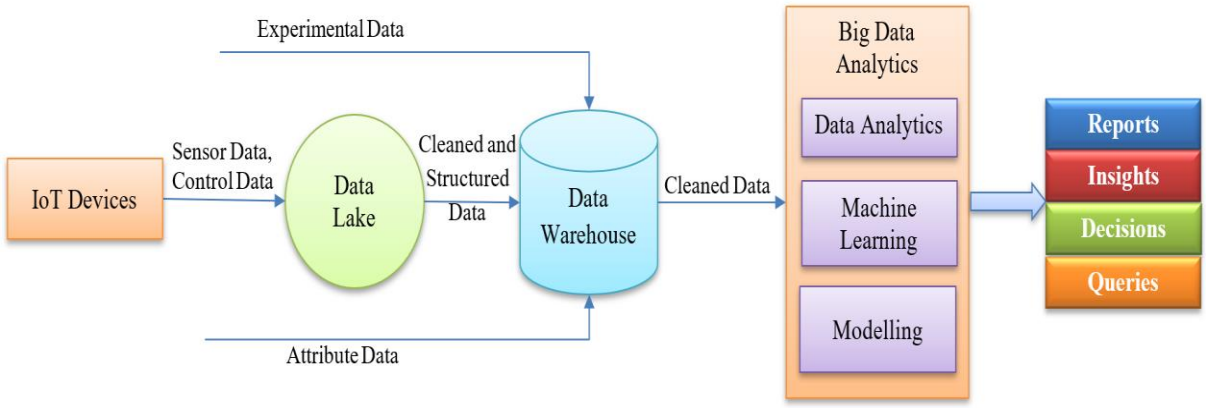

Figure 1: IoT generated Big Data Processing

Big data is more about collecting and accumulating huge data for analysis afterward, whereas IoT is about simultaneously collecting and processing data to make real-time decisions. IoT data is different from as it exhibits the following characteristics: large-scale live streaming data, support data heterogeneity, and correlation between time and space.

Big Data Analytics

Big Data Analytics is the process of examining large data sets containing a variety of data types to uncover hidden patterns, market trends, unknown correlations and other useful information. Big data analytics in IoT involves streaming and storing a large volume of data, and providing real-time services and monitoring. It uses advanced analytic techniques against very large and diverse data sets from different sources and in different sizes, with the additional benefits of speed and efficiency. Big data analytics platform take unstructured data collected by IoT devices and organize these into structure data set to provide information to the companies for optimizing their business problems. Thus, it assists business and other organizations to achieve an improved understanding of data, which enables making

efficient and well-informed decisions. Big data analytics also gives analytics professionals, such as data scientists and predictive modelers, the ability to analyze big data from multiple and varied sources, including transactional data and other structured data. The Big data analytics helps analyzing data in following ways (Figure 2):

Descriptive Analytics

Descriptive analytics is useful for uncovering patterns from past data. It provides insight into what has happened historically. Summary statistics, segmentation and clustering are some techniques for descriptive analysis. In case of time-sensitive data, it becomes easier to perform descriptive analysis with more recent data.

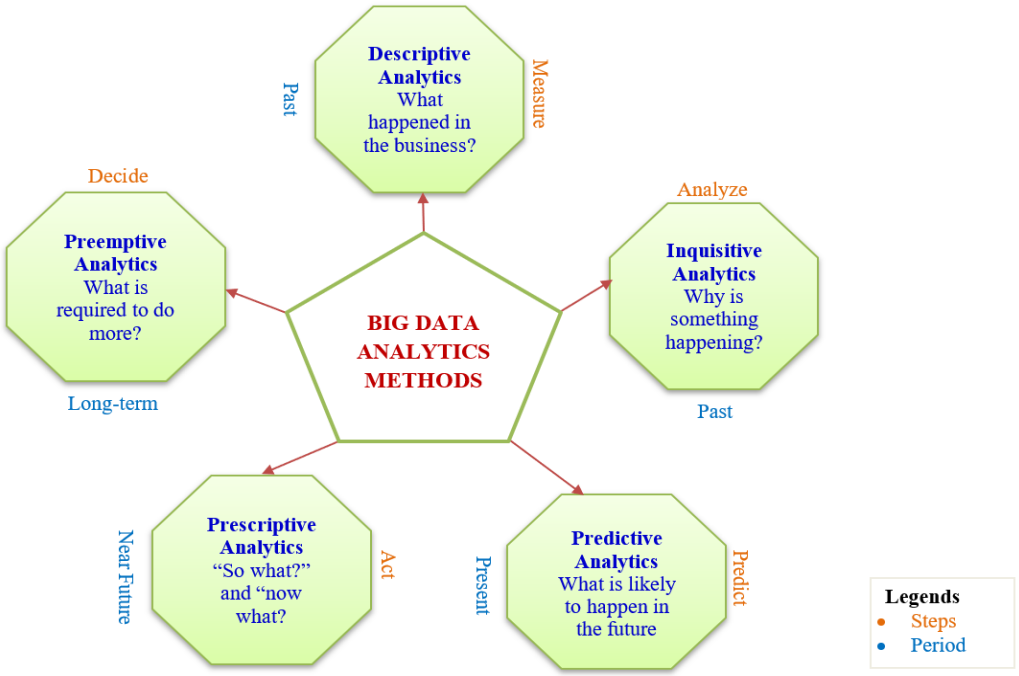

Figure 2: Big Data Analytics Methods

Inquisitive analytics

This helps on probing data to certify/reject business propositions, for example, analytical drill downs into data, statistical analysis, and factor analysis**.** It ascertains historical data against other data to reveal route-cause of the problem. It is used to analyze data with a past event or situation. Data mining, correlation and drill-down are some of the techniques for this type of analysis.

Predictive Analytics

Predictive Analytics involves analysis of past data patterns and trends to accurately forecast the future outcomes. It helps in determining realistic goals for the company and its effective execution and moderating expectations, by manipulating the findings of descriptive and inquisitive analytics. Predictive analytics uses models to forecast what might happen in specific scenarios. It focuses on non-discrete predictions of future states, relationship, and patterns. Statistics, modeling, data mining and machine learning are some techniques for predictive analysis.

Prescriptive Analytics

The main objective of prescriptive analytics is to prescribe the best possible action for situations as it is unfolding. It helps to determine the best solution among a variety of choices, given the known parameters and suggests options for how to take advantage of a future opportunity or mitigate a future risk. What-if analysis can be used for such type of analysis

Pre-emptive Analytics

Pre-emptive analytics helps in taking preventive actions. It has the capacity to take precautionary actions on events that may undesirably influence the organizational performance, for example, identifying the possible perils and recommending mitigating strategies far ahead in time.

These approaches are used to demystify IoT generated big data and convert it into actionable insights for gaining a competitive edge, and making proactive business decisions. Methods/approaches to analyze such data have been discussed in different domains such as banking by Sun, Morris, Xu, Zhu & Xie (2014) and Srivastava & Gopalkrishnan (2015); healthcare by Bates, Saria, Ohno-Machado, Shah & Escobar (2014), Ali et al. (2021) and Stewart & Velupillai (2021); Manufacturing by Clarke (2020), Cui, Kara & Chan (2020) and Kovacova, Segers, Tumpach & Michalkova (2020); Production and Operations Management by Choi, Wallace & Wang (2018); security by Cardenas, Manadhata & Rajan (2013); and supply chain management by Vieira, Dias, Santos, Pereira & Oliveira (2020).

Big data analytics is emerging as a key to analyzing IoT generated data from *connected devices* which helps improving decision making process. The techniques for big data analytics encompass a wide range of mathematical, statistical, and modeling techniques. Statistical modeling is utilized to discover knowledge and wisdom through descriptive analysis that can support decision making. Data mining employs advanced statistical tools to analyze the big data available through data warehouses and

other sources to identify possible relationships, patterns and anomalies and discover information or knowledge for business decision making. Visualization technologies make knowledge patterns and information for decision making in a form of figure, table or multimedia. Thus, big data analytics can facilitate business decision making through analyzing current problems and future trends, and optimizing business processes to enhance organizational performance.

Big Data and Decision Making

In the current era, globalization is blurring the boundaries among nations, thus business executives are challenged with high expectations from customers, high competition, rising costs of labor and materials and shorter product lifecycles. In this volatile environment, firms need to continuously scan for risks and opportunities, and make business decisions quickly based on available data (Schrage, 2016). Big data helps in making better predictions and smarter decisions in the areas such as transactional data, social media data, supply chain big data (Clarke, 2020; Wang & Wang, 2020; Zhan & Tan, 2020). Thus, there is a need to integrate big data technologies with decision making techniques. Horita, de Albuquerque, Marchezini & Mendiondo (2017), Janssen, van der Voort & Wahyudi (2017), Agarwal et al. (2018), Duan, Edwards & Dwivedi (2019), Shamim, Zeng, Shariq & Khan (2019) and Kauffmann et al. (2020) have discussed methods to integrate big data technologies with decision making techniques.

A decision support system (DSS) is an interactive information system that analyzes large volumes of data for informing business decisions. It supports the management, operations, and planning levels of an organization in making better decisions by assessing the significance of uncertainties and the tradeoffs involved in making decision. A DSS leverages a combination of data, documents, personal knowledge, and/or business models to help users make decisions. The data sources used by a DSS may include relational data warehouses, data sources, sales projections, electronic health records (EHRs), cubes and so on.

Categories of decision support systems are:

- **Data-driven DSS:** These systems focus on the methods to access to and manipulate large databases of structured data, usually a time-series of internal company data and sometimes external data.

- **Model-driven DSS:** These systems focus on development and manipulation of a model. Using statistical and analytical tools, as well as OLAP systems. The DSS uses data and parameters provided by decision-makers and are usually not data-intensive.
- **Knowledge-driven DSS:** These systems recommend actions to managers and provide specialized problem-solving expertise based on a particular domain. These systems use data mining to sift through databases for producing data content relationships.
- **Document-driven DSS:** These systems integrate storage and processing technologies for document retrieval and analysis. A search engine is an example of document-driven DSS.
- **Communication-driven and Group DSS:** Communication-driven system focuses on communication, coordination and collaboration to help decision makers working on a shared task. Group DSS focuses on supporting groups of decision makers for analyzing problems and performing group decision-making tasks.

Decision support systems consist of three main components:

1. **The database:** The database is designed through a variety of sources, including internal data the organization, data generated by applications, and external data purchased from third parties or mined from the Internet. The size of the DSS database varies based on need, from a small and standalone system to a large data warehouse.
2. **The Model base:** A model, including decision context and user criteria, is built. For to support decision making and number of models depend on the purpose of the DSS. Commonly used models include:

- **Statistical models**, used to establish relationships between events and factors related to that event.
- **Sensitivity analysis models**, used for what-if analysis.
- **Optimization models**, used to find the optimum value for a target variable in relation to other variables.
- **Forecasting models**, include regression models, time series analysis, and other models used to analyze business conditions and make plans.
- **Backward analysis sensitivity models**, set a target value for a particular variable and then determine the values other variables need to hit to meet that target value.

3. **The user interface:** This component includes dashboards and other user interfaces that allow users to interact with and view results.

LITERATURE REVIEW

With the evolution of technology and the increased multitudes of data flowing in and out of organizations daily, there has become a need for faster and more efficient ways of analyzing such data. These data sets cannot be analyzed with traditional data management and analysis techniques and infrastructures. Therefore, researchers are developing new methods for big data analytics, as well as architectures for storing and managing such data.

IoT generated Big Data

IoT an evolving paradigm generates enormous amount of data, and it becomes challenging to manage and control this big data. Kakiz, Öztürk & Çavdar (2017) proposed an IoT architecture using Software Defined Networking to overcome these challenges. Rather than evaluating the sensor values in the application layer, these were analyzed in lower layers, especially in gateway layer, before being sent to the internet. This architecture enabled handling the big data problem because only useful and helpful packets were sent to the internet. IoT provides the tools for developing global data-driven ecosystem that can be accessible to businesses and people. These models assume the existence and development of the necessary IoT measurement and control instruments, communications infrastructure, and easy access to the data collected and information generated by any party. Lokshina, Durkin & Lanting (2017) presented an overview of business models and opportunities for third-party data analysis services and discussed methods to access information generated by third parties in relation to big data techniques. Novel advances in Information and Communication Technology have revolutionized various areas, however their utilization usually possesses a negative influence on the human health and the environment. Green IoT takes one of the most important roles for creating; a green and sustainable place for living. Big data analysis enabled generating valuable insights from Green IoT generated data. The obtained knowledge enables smart city services like decision-making and forecasting to be performed easily. Maksimovic (2017) summarized the role of big data' achievements in creating cities to enhance quality of life by reducing pollution and efficiently utilizing goods. In the IoT era today, security has become a fundamental factor of most applications, and mechanisms need to be designed to protect

communications enabled by such technologies. Puthal, Ranjan, Nepal & Chen (2017) analyzed mechanisms for securing the IoT and big data, as well as security threats in the domain. The IoT architecture was divided into several layers to define properties, security issues and related works to solve the security concerns. Rathore, Paul, Ahmad & Jeon (2017) presented an IoT-based system for Super City planning using Big Data Analytics. Authors developed a system that includes different IoT-based systems like smart home, weather and water system, vehicular networking and surveillance objects for generating data. An architecture is proposed that includes four tiers i.e., (Bottom Tier-1, Intermediate Tier-1, Intermediate Tier 2, and Top Tier) to handle data generation and collections, communication, data administration and processing, and data interpretation. The system is implemented using Hadoop ecosystem with MapReduce programming. Most of the existing IoT devices do not cooperate with one another by using the same medium of communication. Rathore, Paul, Ahmad, Anisetti & Jeon (2017) developed an intelligent healthcare system to show IoT-based collaborative contextual Big Data sharing among the healthcare devices. The network architecture of the developed system enables to collect data generated by millions of connected devices. This data is then used to identify serious and abnormal health conditions by forwarding the data to intelligent building. The components of intelligent building includes a Big Data collection unit for collecting and filtering data, and for load balancing; a Hadoop processing unit comprised of Hadoop Distributed File System and MapReduce; and an analysis and decision unit. These units are equipped with a medical expert system which reads the sensor data and performs actions in the case of an emergency.

Liu (2018) detailed the main aspects including challenges of IoT that influence data processing, and reviewed the ways to enhance data veracity and mine uncertain data streams. Recommendation related to the heterogeneous and distributed IoT data, autonomous decision-making, domain-optimized methodologies, and secure data management were also presented. Syafrudin, Alfian, Fitriyani & Rhee (2018) developed a real-time monitoring system for manufacturing process by utilizing IoT sensors (to collect humidity, temperature, gyroscope, and accelerometer data) big data processing, and a hybrid prediction model. Apache Kafka was used as a message queue, Storm was used as a real-time processing engine and MongoDB was used to store the sensor data. DB-SCAN (Density-Based Spatial Clustering of Applications with Noise based outlier detection) and Random Forest classification were utilized for

removing outlier sensor data and for providing fault detection during the manufacturing process. The proposed model was evaluated and tested at an automotive manufacturing assembly line in Korea. Yang, Zheng, Guo, Liu, & Chang (2018) proposed a privacy-preserving e-health system, which combines IoT, big data and cloud storage. Physiological data of the patient was monitored through medical IoT network and was aggregated to EHR. This medical big data is outsourced to cloud platform. The IoT messages were encrypted and transmitted to the patient in a fine-grained access control manner. To enable flexible access, without privacy leakage, a keyword match-based update mechanism was designed. Wang, Bhuiyan, Wang, Rahman & Cao (2018) developed a cloud-based health monitoring system, the Big Reduce, with the IoT framework. Real-time data processing was integrated with intelligent decision-making capabilities within the developed system.

Alsaig, Alagar, Chammaa, & Shiri (2019) developed a data-centric approach to IoT in conceptualizing the objects from a service-oriented perspective and investigated the ways to identify, integrate, and manage big data. Authors suggested that the data-centric approach better supports data management with complexities inherent in IoT-generated big data. Reasoning and scalable query processing techniques were also supported by the system in developing smart city applications. Limkar & Jha (2019) developed a method for parallel construction of R Tree and its variants, by using Apache Spark based on the IoT Zetta platform to index real time geospatial data for continuously updating the position of aircraft in real time. R tree was indexed in it to fetch real time results of the spatial query. Taher, Mallat, Agoulmine & El-Mawass (2019) developed an IoT-Cloud based framework for real-time and batch processing of Big Data in the healthcare domain. The framework solution was implemented using Amazon Web Services and Raspberry pi was used as an IoT device to generate data in real time. The performance of the system was tested in terms of response time by varying the velocity and volume of the analyzed data.

Natural as well as human-instigated disasters result in infrastructural damages, distresses, revenue losses, and injuries in addition to huge death roll. Anbarasan (2020) developed methods for the flood disaster detection based on IoT, Big Data, and convolutional deep neural network. The input data was taken from the flood Big Data. The data was pre-processed using missing value imputation and normalization function and the rules were generated by using a combination of attributes method.

Finally, the generated rules were provided as the input to the convolutional deep neural network classifier to classifies them as chances for the occurrence and no chances for the occurrence of flood. Atiewi (2020) presented a cloud-enabled IoT environment supported by multifactor authentication and lightweight cryptography encryption schemes to protect big data system. The hybrid cloud environment combination of private and public cloud enabled protecting organizations' data in a highly secure manner. The IoT devices were divided into sensitive and nonsensitive devices. Authentication levels for accessing the stored data were provided as read the file, download the file, and download file from the hybrid cloud. Balakrishna & Thirumaran (2020) presented an overview of IoT, Cloud Computing and U-health care systems. Semantic annotations were added to achieve semantic interoperability using use-case approaches. SPARQL queries and resource description framework were applied to extract and visualize the health care data. A resource description framework graph was generated which was semantically interoperable and intercommunicated between the IoT-based smart home resources (implemented through Restlet framework tool). As IoT plays an important role in developing new source dataset, a seamless authentication protocol was developed by Fadi & Deebak (2020) to merge data inference. Comparative analysis was performed to show that the proposed protocol consumes less computation and communication overhead as compared to other authentication schemes. IoT generated big data brings new opportunities to monitor agricultural and food processes. Misra et al. (2020) presented the role of IoT, big data, and artificial intelligence in shaping the future of agri-food systems. Authors discussed the role of big data analysis based on IoT in drone-based crop imaging, greenhouse monitoring, food safety, and food quality assessment. Sestino, Prete, Piper & Guido (2020) investigated the role of the IoT and Big Data in managing their digital transformation of their businesses and argued that this has resulted in a mass of disorganized knowledge. Upon analyzing both the IoT and Big Data, authors demonstrated that these are predominantly reengineering factors for business processes. Authors highlighted suggestions for managers/marketers to show how digital transformation enabled by IoT generated big data can have positive impacts on various aspects of business.

Bamakan, Faregh & ZareRavasan (2021) developed a distributed learning framework to evaluate service supply chain performance based on blockchain and ANFIS (adaptive network-based fuzzy inference systems) techniques. Authors developed a smart learning model to deal with the

uncertainty conditions by a combing neural network and fuzzy logic. A distributed blockchain-based framework was also presented to ensure security in the supply chain. Colossal development in the field of artificial intelligence, connectivity, machine learning and IoT has made it possible to acquire an understanding of social as well as machine behavior to a great extent. Kashyap (2021) explored the application of IoT generated big data for environmental sustainability in developing countries, particularly India. The study also proposed a roadmap for the same. Rashid, Singh, Goyal, Parah, & Wani (2021) aimed at solving the challenge of real time processing of sensor based data and processing it for classification using hybrid machine learning classifier. The IoT based sensor data is stored on Hadoop Distributed File System which is processed on Big Data processing pipeline. The outliers in stored data were removed with the help of density-based spatial clustering of applications with noise and then unwanted features were eliminated using Recursive Feature Elimination with Cross Validation. Random Forest ensemble classifier was used for classifying classes in medical data.

IoT generated Big Data Analytics

In order to efficiently analyze the enormous amount of data generated by connected devices, researchers have developed various models/tools for determining customer preferences, future trends and other information. Al-Ali, Zualkernan, Rashid, Gupta & AliKarar (2017) presented an Energy Management System for smart homes. Each home device of the smart home was interfaced with an IoT object for data acquisition. Energy consumption data was also collected from each device of the smart home and was transmitted to a centralized server for processing. This data, collected from all residential areas, was gathered in a server as Big Data. Big Data analytics was then performed to manage energy consumption and meet consumer demand. Spiliopoulos, Zissis & Chatzikokolakis (2017) described a four-step approach, based on the MapReduce paradigm, to transform spatiotemporal data into information for understanding the patterns of global trade by adopting distributed processing approaches.

Kong et al. (2018) proposed a robust incremental on-line feature extraction method based on Principal Component Analysis, and Robust Incremental Principal Component Analysis. Nair, Shetty & Shetty (2018) developed a real time remote health status prediction system using Apache Spark, deployed it in the cloud, and applied machine learning model to predict user's health status which is then directly messaged to the patient for taking appropriate action. ur Rehman et al. (2018) introduced a

concentric computing model paradigm composed of sensing systems, outer and inner gateway processors, and outer and inner central processors for the deployment of big data analytics applications in Industrial IoT. Authors also identified and discussed indispensable challenges in employing concentric computing model in the Industrial IoT paradigm. Based on practice-based view, Wang, Wang, & Cegielski (2018) developed a big data analytics-enabled transformation model to uncover the causal relationships among big data analytics capabilities, benefit dimensions, IT-enabled transformation practices, and business values.

A smart transportation system was developed by Babar & Arif (2019) by performing big data analytics to achieve real time processing. The system followed a three-phased scheme to organize and manage big data, and its real-time processing and service management. The developed scheme utilizes Spark over single node Hadoop setup with various input libraries. Bhattarai (2019) reviewed the application of big data analytics in power grids and identified challenges and opportunities from industry and research perspectives. Authors also presented visions to integrate big data analytics into power system planning and operational frameworks. With an exponential growth of raw data generated from different IoT devices, data analytics is becoming a challenging issue in terms of heterogeneity of the data generated. Kumari, Tanwar, Tyagi & Kumar (2019) presented an overview on handling secured streaming data generated by IoT devices during big data processing. The architectural details and security approaches required in each phase of big data processing was also explored in detail. Popchev, Orozova & Stoyanov (2019) developed a conceptual model for virtualizing education in the IoT paradigm. The impact of IoT and big data analytics in the field of virtual education space was also presented. Shah et al. (2019) proposed reference architecture of a disaster resilient smart city by integrating IoT and Big Data Analytics technologies. The architecture helped in determining solutions for disaster management activities in smart city incentives. Hadoop Ecosystem and Spark were used in combination to develop disaster resilient smart city environment for supporting both real-time and offline analysis. Smart buildings, city pollution, traffic simulator, and twitter datasets were utilized for detecting and generating alerts in case of fire in a building, determining pollution level in the city, finding emergency evacuation path, and collecting information about earthquakes and tsunamis. Toapanta, Romero, Saltos & Gallegos (2019) analyzed a part of the global ecosystem with theoretical

and applied security models. The goal was to define a prototype of security model for IoT with Big Data for tourism management in cities of Ecuador. Connected appliances and devices inside a smart home produce significant amount of data about consumers and their daily activities. Yassine, Singh, Hossain & Muhammad (2019) performed analytics for IoT generated data from smart homes by using fog nodes to allow data-driven services. Authors also addressed the challenges of complexities and resource demands for data processing, storage and classification analysis.

Big data analytics and IoT tools are considered crucial investments for firms to distinguish themselves among competitors. Côrte-Real, Ruivo & Oliveira (2020) reviewed the strategic management perspective of big data analytics and IoT capabilities that can create significant value in business processes if supported by a good level of data quality. Responses were collected from 618 European and American firms that use IoT and Big data analytics applications and were analyzed using Partial least squares. Entezami, Sarmadi, Behkamal & Mariani (2020) suggested a approach based on autoregressive moving average model for feature extraction. Authors developed a hybrid divergence-based method for feature classification. Data relevant to a cable-stayed bridge was generated to assess the effectiveness and efficiency of the developed method. Li (2020) developed IoT based Smart and real time healthcare information processing system for continuous and real-time data monitoring and processing. The Author also discussed the plan for evaluating its efficacy in terms of storage and transmission costs, accuracy, sensitivity and other parameters. Liu, Feng, Lin, Wu & Guo (2020) developed Cloud Laundry, an IoT based E-commerce model, for mass scale laundry services by using intelligent logistics management, big data analytics, and machine learning techniques. Best transportation path was computed using GPS and real-time update and logistic terminals were updated and rerouted simultaneously. Using the user's specifications, the system dynamically provides best laundry solutions based on the current state spaces of the laundry terminals and offers the customers with convenient and transparent laundry services. Cloud laundry also uses big data models to maintain customers' security needs. Özemre & Kabadurmus, (2020) employed Random Forest and Artificial Neural Networks to forecast export volumes using an open trade data. The forecasted values were included in the Boston consulting group matrix to conduct strategic market analysis.

Sasubilli Kumar & Dutt (2020) collected and analyzed the tribal people health information and developed a system to perform the treatment with fewer prices. Authors also predicted the disease and the drug cases using past data for various seasons and various diseases. Yasmin, Tatoglu, Kilic, Zaim, & Delen (2020) explored three multicriteria decision-making methodology tools: intuitionistic fuzzy decision-making trial and evolution laboratory, analytic network process and simple additive weighting, to assess the relative importance of big data analytics capabilities and the relationship of these capabilities with the firm performance.

Chen, Sun, Chang & Lin (2021) developed a gradient boosting partitioned regression tree model for forecasting travel time based on big data collected from the industrial IoT infrastructure. The method separated the global regression tree model based on the gradient boosting decision tree into several partitions to capture the time-varying features simultaneously. Several analytical properties were also illustrated. The model was applied to real traffic IoT data. Kanaga & Jacob (2021) presented a framework for smart waste management based on IoT using Big Data Analytics. Data was generated from IoT-based smart dustbins, and then was aggregated. Cleaned optimized schedule was generated for waste collection based on the sensor data. Implementation was done using Hadoop framework and IBM InfoSphere Streams. A framework consisting of data collection module, data storage module, Hadoop/Spark cluster, and Google Cloud was developed using decision tree and support vector machine techniques by Qaffas, Hoque & Almazmomi (2021) for predicting hypertension. The study was performed over 140 participants. The results show that age and diabetes play a significant role in diagnosing hypertension in older people. Unal et al. (2021) suggested to use Blockchain for mitigating attacks on Federated Learning algorithms operating in IoT systems. Blockchain and Federated Learning were integrated for securing the trained models' integrity to prevent model poisoning attacks. To provide the security of user data and the trained models, author utilized fuzzy hashing to detect variations and anomalies in Federated Learning -trained models. Singh, Kaur, Batth, Kaur & Gianini (2021) developed a model by applying beetle swarm optimization and adaptive neuro-fuzzy inference system for heart disease and multi-disease diagnosis. The main components of the analytics pipeline were the modified crow search algorithm which was used for feature extraction, and an ANFIS

classification model whose parameters were optimized by means of a beetle swarm optimization algorithm.

IoT generated big data and decision making

In order to support the management, operations, and planning levels of an organization, DSSs play a vital role. Integrating IoT generate big data enables smarter decisions for real world problems. Bumblauskas, Gemmill, Igou & Anzengruber (2017) outlined the architectural design and the conceptual framework for a Smart Maintenance DSS based on corporate data of Fortune 500 Company. The System is capable of providing end users with recommendations to improve asset lifecycles. A cost minimization algorithm was developed to analyze the large industry service and warranty data sets. Two analytical decision models were developed for an electrical circuit breaker maintenance problem. Big Data has a great impact on the cities' morphology and urban planning. Guido, Rogano, Vitale, Astarita & Festa (2017) presented a DSS framework for generating travel strategies by expanding a large amount of transportation systems data. A framework was developed for collecting, aggregating, and disseminating data coming from different sources. Data mining approaches were applied to analyze open Big Data and unstructured freely available Big Data. Experimental data was collected from both a sample of smartphones and automatic vehicle location systems in the case study area. Kaplunovich & Yesha (2017) developed a recommendation system to choose the best instance for a dataset and a machine learning algorithm to optimize time and money spent. Algorithms for big data sets on miscellaneous Amazon Web Services instances and collecting the results in the DynamoDB database, machine learning models were trained to predict time and cost using various regression machine learning methods. Voss, Sebastian & Pahl (2017) discussed the role of intelligent decision support and big data for logistics and supply chain management by describing theoretical developments.

Boulila, Farah & Hussain (2018) developed a tool for interpreting and analyzing Remote Sensing big data to gather knowledge for decision support in Remote Sensing fields. It helped users easily make decisions by providing descriptive, predictive and prescriptive analytics. The design and development of framework was based on: Remote Sensing data acquisition, modeling, and analysis and interpretation. The models were developed using clustering, decision tree and association rules. A DSS was developed by Jamshidi (2018) for condition-based maintenance of rails. The methodology

considered both the actual conditions of the rails (using axle box acceleration measurements and rail video images) and the prior knowledge of the railway track and a provided an integrated estimation of the rail health conditions and grind planning of the segment that are prone to critical conditions. Kourtit & Nijkamp (2018) addressed the strategic governance challenges of modern smart cities from viewpoint of big data management. In order to enhance the competitive performance of the smart urban cities, a framework was also developed to deal with multivariate big data in the decision support context. The developed dashboard system functioned as an interactive navigation tool. Lee, Aydin, Choi, Lekhavat & Irani (2018) estimated a fuel consumption function to optimize the speed problem by considering weather data. Copernicus data set was used as the source of big data. Data mining technique was applied to identify the impact of weather conditions based on a given voyage route. Particle swarm optimization was applied to find pareto optimal solutions that minimize fuel consumption and maximize service level agreement. Stein, Meller, & Flath (2018), by augmenting the manufacturing process with highly sensitive sensors, described challenges in developing a predictive DSS. Due to advancements in machine learning techniques and availability of patients' big data, diagnostic decision support in radiology is experiencing rapid transformation. Syeda-Mahmood (2018) described the growth of diagnostic decision support for radiology by highlighting progress in this field from the early days rule-based expert systems to modern era cognitive assistants. Zhao & Ming (2018) utilized big data to develop a DSS for smart campus. The DSS for smart education provided decision for various campus operations like resources utilization, staff management, and teaching evaluation.

Kindle, Badawi, Celi & Sturland (2019) presented an overview of Machine Learning applications in the ICU that may be suitable for tele-ICU adaptation. Polyakova, Loginov, Strelnikov & Usova (2019) developed social network analysis tools to discover the nature of connections between objects and to find solutions for various analytical tasks. In recent years, big data systems are getting increasingly complex and require a deep domain specific knowledge. Volk, Bosse, Bischoff, & Turowski (2019) utilized a design science research methodology to develop a DSS. Big data reference architectures were identified and compared with each other. Use of Analytic Hierarchy Process was suggested as the main artefact and was evaluated on a real-world use-case.

With the rapid increase in technology, the healthcare sector is approaching towards digital solutions for clinical evidence related to various diseases. Clim, Zota, Constantinescu & Ilie-Nemedi (2020) developed a theoretical framework for a Clinical DSS by using Software as Medical Device regime. The developed model followed a layered architecture where each layer acquired and processed data obtained through EHR in hospitals. The layers of the model involve sensors for measuring data, pre-processing and transmission to the cloud servers for storage, data reduction, and feature extraction. The model assisted doctors in deciding on diagnosis and treatment of diseases accurately. Usage of big data and blockchain technologies has given the new paradigm shift in sustainable agriculture research. Dakshayini & Prabhu (2020) presented a technology-based idea to address the problems of agriculture effectively. Author concluded that the model improved the quality of the agricultural system by minimizing the gap among demand and supply of food crops required by the society from the farmer's end. An intelligent DSS based on data mining technology was developed by Guo, Wang, Xu, & Wu (2020) to support the decision-makers in making intelligent decisions. The results shown that data mining technology can analyze the statistical data from multiple angles and perspectives by modeling, classifying, and clustering a large amount of data, as well as discovering the correlations between the data. In order to protect human lives and infrastructure and to minimize the risk of damage, it is important to predict to natural disasters. Jung, Tran Tuan, Dai Tran, Park, & Park (2020) developed an intelligent disaster management system framework particularly for cold/heat waves and wildfires. The DSS used big data collected from artificial intelligence algorithms and open APIs for making accurate and faster decisions. Convolutional neural network was also developed to detect fire in surveillance video, which can be used for automatic fire detection and provide an appropriate response. A huge data set of X-ray images from patients with common bacterial pneumonia, confirmed Covid-19 disease, and normal healthy cases were utilized by Karthikeyan, Varde & Wang (2020) for Artificial Intelligence based decision support in detecting the Coronavirus disease. The transfer learning approach enabled to learn from a smaller set of samples in a problem and transfer the discovered knowledge to a larger data set. Transfer learning was considered using three different models that are pre-trained on several images from the ImageNet source. Big data and artificial intelligence are most trending tools for innovation and predictive analytics in healthcare. López-Martínez, Núñez-Valdez, García-Díaz, & Bursac (2020)

discussed the role of intelligent big data analytics to improve population health management, value-based care, and new upcoming challenges in healthcare. Design of a digital health platform was presented to integrate operational, clinical, and business data repositories with advanced analytics for improving the health management decision-making process. Lu, Liu, Song & Zhang (2020) discussed the basic framework and prevailing techniques in streaming type big data and concept drift for Data-driven decision-making. A technical framework for real-time data-driven decision-making was established, methodologies for detecting concept drift and supporting data-driven decision-making were presented, and the characteristics of high-volume streaming data were detailed. A vital division in any organization is the department of human resources, which is in charge of the recruitment and personnel selection procedures. Maghsoodi, Riahi, Herrera-Viedma & Zavadskas (2020) extended the CLUS-MCDA approach (CLUSter analysis for improving Multiple Criteria Decision Analysis) and integrated it with the Best–Worst Method to solve multi-scenario big data decision-making problems., Multiple personnel selection and risk assessment problems were investigated for validating the reliability and practicality of the approach. Data-driven decision-making approach was used for multi-scenario parallel decision making. Massaro, Ricci, Selicato, Raminelli & Galiano (2020) developed a DSS based on health wearable sensors and AI algorithms for predicting the health status. An innovative architecture of a prototype platform related to a research industry project was presented. The developed prototype platform enabled controlling different wearable sensors storing data into a Cassandra Big Data system. Long short-term memory and support vector machine algorithms were utilized to perform prediction based on experimental datasets.

Big Data frameworks were used by Sbai & Krichen (2020) for solving the Dynamic Vehicle Routing Problem to suggest optimized route. A parallel Spark Genetic Algorithm was developed to take the advantage of Spark's in-memory computing ability and genetic algorithm's iterations operations. Parallel operations were used for fitness evaluation and genetic operations, based on the parallel Spark Genetic Algorithm. Decision-making in flood risk management increasingly depends on access to data. Towe et al. (2020) examined architecture and associated techniques of data-driven approach to support decision making in flood risk management. The benefit of using cloud-based data hypercube was discussed to reach the desired level of integration of the highly complex data.

A review of DSS and artificial intelligence in forecasting crop disease was presented by Fenu & Malloci (2021). The authors described the DSS LANDS for monitoring the main crop productions in Sardinia. A feed-forward neural network was implemented to investigate if weather data provided by regional stations could be used to predict a disease risk index using an Artificial Intelligence technique. Hamilton (2021) synthesized and analyzed existing evidence on real-time big data analytics, sustainable Industry 4.0 wireless networks, and IoT based DSSs in cyber-physical smart manufacturing. Using the data from different sources, the author performed analyses to make estimates for big data-driven manufacturing. Descriptive statistics of compiled data from the completed surveys were calculated when appropriate. In order to improve the intelligence and real-time performance of enterprise financial decision-making, ID3 algorithm and association rule mining algorithm were adopted by Si & Shi (2021) to analyze and optimize the financial DSS. Wang, Lin, Chu, Tao & Ling (2021) incorporated Remote Sensing technology with management strategies and proposed a Remote Sensing Knowledge-Based DSS framework. It helped in conducting strategic thinking that is needed for keeping the inventory and monitoring task running by obtaining knowledge and decision support. Watkins (2021) evaluated and analyzed real-time big data analytics, smart industrial value creation, and robotic wireless sensor networks in IoT-based DSSs. Analyses was performed regarding product decision-making information systems in data-driven sustainable smart manufacturing. Descriptive statistics of compiled data from the completed surveys were calculated when appropriate.

Motivation

Research gaps revealed from the literature review

- In order to make sense of all heterogeneous and fragmented information, the need is to research how big data analytics can be integrated into the decision-making process.
- Real-time monitoring systems based on IoT devices, big data processing and hybrid prediction models are very few.

Developing Big Data applications has become increasingly important in the last few years. In fact, several organizations from different sectors depend increasingly on knowledge extracted from huge volumes of data. Solutions need to be studied and provided to extract valuable knowledge from these datasets. In order to make sense of all this heterogeneous and fragmented information, need is to research

how big data analytics can be integrated into the decision-making process. This research study is an effort in this direction.

PROPOSED WORK

The focus of this study is development of a hybrid decision support framework for IoT based big data environment.

Objectives of the proposed study

- Preparation of data lake and its transformation into an ordered data repository,
- Development of a hybrid methodology by integrating big data analytics with decision making technique(s),
- Development of a hybrid decision support system, and
- Implementation of the proposed system for a real-world problem.

Methods to achieve the proposed objectives

O1: Preparation of data lake and its transformation into an ordered data repository *(may be done tentatively within first 6 months)*

After deciding upon the identified requirements (through discussions or surveys) for the proposed study, diverse data sources will be identified. All required data will then be gathered from the identified internal and/or external data sources and stored in data lakes. This data lake will yield deep and actionable insights to create business value. Methods will be developed to preprocess this data to store in an ordered data repository. Methods will also be developed to generate actionable insights from this repository.

As the semi-structured and unstructured data may not fit well in traditional data warehouses based on relational databases, the IoT generated big data will be collected, processed and analyzed using an open-source software framework including Hadoop and related tools such as YARN/MapReduce, and NoSQL databases.

The types of data collected by the IoT devices will include sensor data, location data, status data and spatial data depending upon the implementation area of the proposed study.

O2: Development of a hybrid methodology by integrating big data analytics with decision making technique(s) *(may be done tentatively within next 6 months)*

The proposed system will map big data analytics techniques to different phases of decision-making process. The proposed system will be a learning-based system that will be based on ensemble techniques in machine learning. These techniques will be used to create multiple models and then combine them to produce improved results.

The process will contain following phases (Figure 3):

a. Data Acquisition

This phase will include identifying the data sources (say from traditional data sources such as customer data, multimedia data and so on depending upon the application domain of the study area). The required data will be acquired from internal data-sources (generally structured data) or external data-sources (generally unstructured data) for processing. The collected, stored, and integrated relevant data will be combined from the different data sources to form data lakes.

b. Data Structuring

In order to give efficient decisions, data preparation plays an important role. This phase will require cleaning the impure data so that it can be used for further analysis. Data warehouse will be developed to store multi-dimensional data for query and analysis purpose. This phase will also deal with filtering of data by applying correct metadata and processes. Multiple data sources will be integrated and transformed to add meaning to the data. The main goal here will be extracting relevant information for decision-making from Big Data.

c. Intelligence Phase: The primary goal of this phase is to identify pertinent information by exploiting the generated big data. This phase will use the cleaned data to apply analytics and predictive models for finding patterns that were not initially known. The level of intelligence applied will depend on the computational capabilities and skill-set available together with the business requirements.

Various analysis techniques like statistical tools data mining will be used to extract knowledge from the data. Hadoop provides framework for developing model from different data sources.

The performance of the developed system will be evaluated by considering parameters like resource consumption, execution time, cost, ease of deployment, data visualization.

d. Design Phase: The main objective of this phase is to identify the possible solutions by developing and analyzing the model representing the problem. The steps followed in this phase will be:

- Selection of appropriate model for data analytics
- Application of the selected model
- Analyzing the model outputs

e. Choice Phase: The main purpose of this phase is to evaluate the effects of the selected model in the design phase. The steps followed in this phase will be:

- Evaluating and prioritizing the solutions and their effects by applying big data visualization tools.
- Selecting the best solution.

f. Implementation Phase: The selected solution will be implemented in this phase. Big data methods and techniques will be used to monitor the decision outputs and to provide feedbacks on the implementation results.

O3: Development of a hybrid decision support system *(may be done tentatively within another next 6 months)*

A hybrid DSS will be developed to make the proposed model workable via a user-friendly interface. Emphasis will be placed on both data driven and model driven approaches of the DSS. The conceptual framework for integrating big data and decision making to develop analytics practices for supporting business decisions is shown in Figure 3. The proposed system will be developed to deal with unstructured data collected by IoT devices and organize information into useful datasets using big-data analytics.

In the current environment, sensors send real-time information to the internet (IoT), where it is compiled and stored into vast data set (big data) and analyzed to fi nd useful patterns (hybrid framework). This tremendous volume of data, generated by IoT devices, demands new architectures and technologies to capture and process this big data for enhanced insights and decision making. The proposed hybrid framework will enable to make sense of the plethora of real-time data and provide helpful insights and decision making.

O4: Implementation of the proposed system for a real-world problem *(may be done tentatively with another next 6 months)*

There are several unique research opportunities in different business, scientific and government domains where huge amount of data is generated. The proposed system with its different components or modules will be developed for any one of the following areas: Tourism, Sports, or Higher Education (taking Dayalbagh Educational Institute as a case study), depending upon the suitability of the availability of the data.

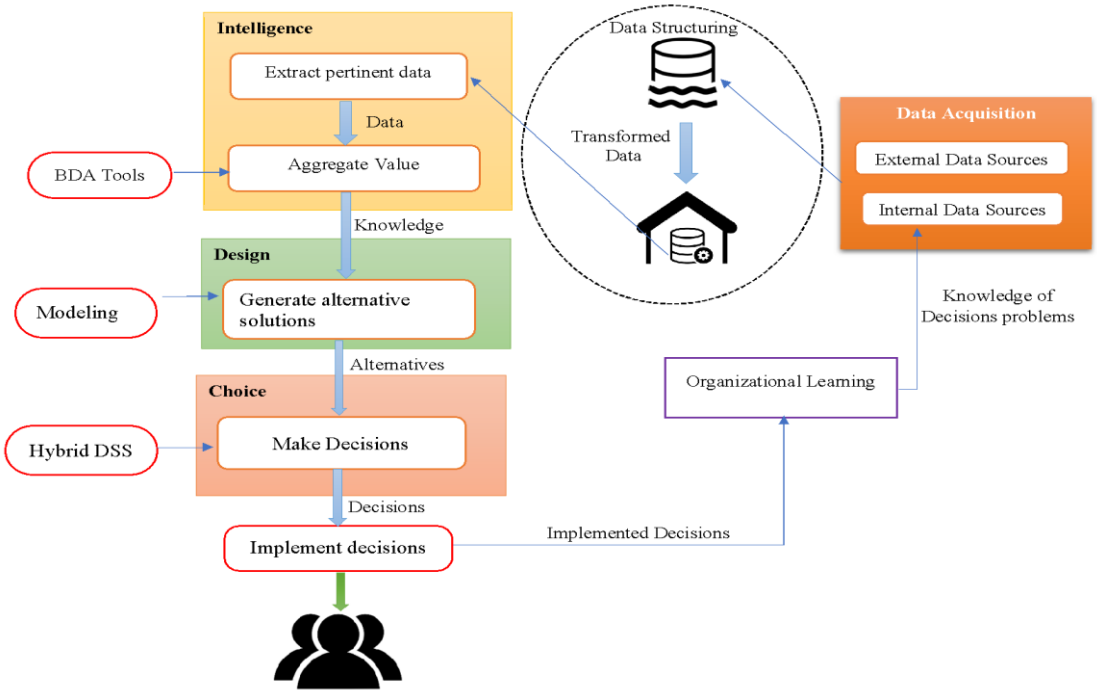

Figure 3: Conceptual framework for the integrated system

Many firms use analytics in almost all aspects of conducting their business to reap the benefits of analytics-based decision making. The proposed system will be developed to improve their analytics capabilities and help business executives to determine competitive intelligence and generate cost and time reduction strategies.

References

1. Agarwal, S., Narayanan, K., Sinha, M., Gupta, R., Eswaran, S., & Mukherjee, T. (2018, July). Decision Support Framework for Big Data Analytics. In *2018 IEEE World Congress on Services (SERVICES)* (pp. 53-54). IEEE.
2. Al-Ali, A. R., Zualkernan, I. A., Rashid, M., Gupta, R., & AliKarar, M. (2017). A smart home energy management system using IoT and big data analytics approach. *IEEE Transactions on Consumer Electronics*, *63*(4), 426-434.

3. Ali, F., El-Sappagh, S., Islam, S. R., Ali, A., Attique, M., Imran, M., & Kwak, K. S. (2021). An intelligent healthcare monitoring framework using wearable sensors and social networking data. *Future Generation Computer Systems*, *114*, 23-43

4. Alsaig, A., Alagar, V., Chammaa, Z., & Shiri, N. (2019). Characterization and efficient management of big data in iot-driven smart city development. *Sensors*, *19*(11), 2430.

5. Alturki, B., Reiff-Marganiec, S., & Perera, C. (2017, October). A hybrid approach for data analytics for internet of things. In *Proceedings of the Seventh International Conference on the Internet of Things* (pp. 1-8).

6. Anbarasan, M., Muthu, B., Sivaparthipan, C. B., Sundarasekar, R., Kadry, S., Krishnamoorthy, S., & Dasel, A. A. (2020). Detection of flood disaster system based on IoT, big data and convolutional deep neural network. *Computer Communications*, *150*, 150-157.

7. Atiewi, S., Al-Rahayfeh, A., Almiani, M., Yussof, S., Alfandi, O., Abugabah, A., & Jararweh, Y. (2020). Scalable and secure big data IoT system based on multifactor Authentication and lightweight cryptography. *IEEE Access*, *8*, 113498-113511.

8. Babar, M., & Arif, F. (2019). Real-time data processing scheme using big data analytics in internet of things based smart transportation environment. *Journal of Ambient Intelligence and Humanized Computing*, *10*(10), 4167-4177.

9. Balakrishna, S., & Thirumaran, M. (2020). Semantic interoperability in IoT and big data for health care: a collaborative approach. In *Handbook of Data Science Approaches for Biomedical Engineering* (pp. 185-220). Academic Press.

10. Bamakan, S. M. H., Faregh, N., & ZareRavasan, A. (2021). Di-ANFIS: an integrated blockchain–IoT–big data-enabled framework for evaluating service supply chain performance. *Journal of Computational Design and Engineering*, *8*(2), 676-690.

11. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Affairs, 33(7), 1123-1131

12. Bhattarai, B. P., Paudyal, S., Luo, Y., Mohanpurkar, M., Cheung, K., Tonkoski, R., ... & Zhang, X. (2019). Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. *IET Smart Grid*, *2*(2), 141-154

13. Boulila, W., Farah, I. R., & Hussain, A. (2018). A novel decision support system for the interpretation of remote sensing big data. *Earth Science Informatics*, *11*(1), 31-45.

14. Bumblauskas, D., Gemmill, D., Igou, A., & Anzengruber, J. (2017). Smart Maintenance Decision Support Systems (SMDSS) based on corporate big data analytics. *Expert systems with applications*, *90*, 303-317.

15. Cardenas, A. A., Manadhata, P. K., & Rajan, S. P. (2013). Big data analytics for security. *IEEE Security & Privacy*, *11*(6), 74-76

16. Chen, H., & Fu, Z. (2015). Hadoop-based healthcare information system design and wireless security communication implementation. *Mobile Information Systems*, *2015*.

17. Chen, Y. T., Sun, E. W., Chang, M. F., & Lin, Y. B. (2021). Pragmatic real-time logistics management with traffic IoT infrastructure: Big data predictive analytics of freight travel time for Logistics 4.0. *International Journal of Production Economics*, *238*, 108157.

18. Choi, T. M., Wallace, S. W., & Wang, Y. (2018). Big data analytics in operations management. Production and Operations Management, 27(10), 1868-1883

19. Clarke, G. (2020). Sensing, Smart, and Sustainable Technologies in Big Data-driven Manufacturing. *Journal of Self-Governance and Management Economics*, *8*(3), 23-29

20. Clim, A., Zota, R., Constantinescu, R., & Ilie-Nemedi, I. (2020). Health services in smart cities: Choosing the big data mining based decision support. *International Journal of Healthcare Management*, *13*(1), 79-87.

21. Côrte-Real, N., Ruivo, P., & Oliveira, T. (2020). Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value?. *Information & Management*, *57*(1), 103141.

22. Cui, Y., Kara, S., & Chan, K. C. (2020). Manufacturing big data ecosystem: A systematic literature review. *Robotics and computer-integrated Manufacturing*, *62*, 101861

23. Dakshayini, M., & Prabhu, B. B. (2020). An effective big data and blockchain (BD-BC) based decision support model for sustainable agriculture system. In *EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing* (pp. 77-86). Springer, Cham.

24. Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. International Journal of Information Management, 48, 63-71

25. Elgendy, N., & Elragal, A. (2016). Big data analytics in support of the decision making process. *Procedia Computer Science*, *100*, 1071-1084.

26. Entezami, A., Sarmadi, H., Behkamal, B., & Mariani, S. (2020). Big data analytics and structural health monitoring: a statistical pattern recognition-based approach. *Sensors*, *20*(8), 2328

27. Fadi, A. T., & Deebak, B. D. (2020). Seamless authentication: For IoT-big data technologies in smart industrial application systems. *IEEE Transactions on Industrial Informatics*, *17*(4), 2919-2927.

28. Fenu, G., & Malloci, F. M. (2021). LANDS DSS: a decision support system for forecasting crop disease in Southern Sardinia. *International Journal of Decision Support System Technology (IJDSST)*, *13*(1), 21-33.

29. Guido, G., Rogano, D., Vitale, A., Astarita, V., & Festa, D. (2017, June). Big data for public transportation: A DSS framework. In *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)* (pp. 872-877). IEEE.

30. Guo, Y., Wang, N., Xu, Z. Y., & Wu, K. (2020). The internet of things-based decision support system for information processing in intelligent manufacturing using data mining technology. *Mechanical Systems and Signal Processing*, *142*, 106630.

31. Hamilton, S. (2021). Real-Time Big Data Analytics, Sustainable Industry 4.0 Wireless Networks, and Internet of Things-based Decision Support Systems in Cyber-Physical Smart Manufacturing. *Economics, Management, and Financial Markets*, *16*(2), 84-94.

32. Harb, H., Mroue, H., Mansour, A., Nasser, A., & Motta Cruz, E. (2020). A Hadoop-Based Platform for Patient Classification and Disease Diagnosis in Healthcare Applications. *Sensors*, *20*(7), 1931

33. Horita, F. E., de Albuquerque, J. P., Marchezini, V., & Mendiondo, E. M. (2017). Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil. *Decision Support Systems*, *97*, 12-22

34. Jamshidi, A., Hajizadeh, S., Su, Z., Naeimi, M., Núnez, A., Dollevoet, R., ... & Li, Z. (2018). A decision support approach for condition-based maintenance of rails based on big data analysis. *Transportation Research Part C: Emerging Technologies*, *95*, 185-206.

35. Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. Journal of Business Research, 70, 338-345

36. Jeble, S., Kumari, S., & Patil, Y. (2017). Role of big data in decision making. *Operations and Supply Chain Management: An International Journal*, *11*(1), 36-44

37. Jung, D., Tran Tuan, V., Dai Tran, Q., Park, M., & Park, S. (2020). Conceptual framework of an intelligent decision support system for smart city disaster management. *Applied Sciences*, *10*(2), 666.

38. Kakiz, M. T., Öztürk, E., & Çavdar, T. (2017, September). A novel SDN-based IoT architecture for big data. In *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)* (pp. 1-5). IEEE.

39. Kanaga, E. G. M., & Jacob, L. R. (2021). Smart Solution for Waste Management: A Coherent Framework Based on IoT and Big Data Analytics. In *Intelligence in Big Data Technologies—Beyond the Hype* (pp. 97-106). Springer, Singapore.

40. Kaplunovich, A., & Yesha, Y. (2017, December). Cloud big data decision support system for machine learning on AWS: Analytics of analytics. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 3508-3516). IEEE.

41. Karthikeyan, D., Varde, A. S., & Wang, W. (2020, December). Transfer learning for decision support in Covid-19 detection from a few images in big data. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 4873-4881). IEEE.

42. Kashyap, A. (2021). A Roadmap for Application of IoT-Generated Big Data in Environmental Sustainability. *Big Data Analytics for Internet of Things*, 197-207.

43. Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, *90*, 523-537

44. Kindle, R. D., Badawi, O., Celi, L. A., & Sturland, S. (2019). Intensive care unit telemedicine in the era of big data, artificial intelligence, and computer clinical decision support systems. *Critical care clinics*, *35*(3), 483-495.

45. Kong, X., Chang, J., Niu, M., Huang, X., Wang, J., & Chang, S. I. (2018). Research on real time feature extraction method for complex manufacturing big data. *The International Journal of Advanced Manufacturing Technology*, *99*(5), 1101-1108

46. Kourtit, K., & Nijkamp, P. (2018). Big data dashboards as smart decision support tools for i-cities–An experiment on Stockholm. *Land use policy*, *71*, 24-35.

47. Kovacova, M., Segers, C., Tumpach, M., & Michalkova, L. (2020). Big Data-driven Smart Manufacturing: Sustainable Production Processes, Real-Time Sensor Networks, and Industrial Value Creation. *Economics, Management & Financial Markets*, *15*(1)

48. Kumari, A., Tanwar, S., Tyagi, S., & Kumar, N. (2019). Verification and validation techniques for streaming big data analytics in internet of things environment. *IET Networks*, *8*(3), 155-163.

49. Lee, H., Aydin, N., Choi, Y., Lekhavat, S., & Irani, Z. (2018). A decision support system for vessel speed decision in maritime logistics using weather archive big data. *Computers & Operations Research*, *98*, 330-342.

50. Li, C. (2020). Information processing in Internet of Things using big data analytics. *Computer Communications*, *160*, 718-729.

51. Limkar, S. V., & Jha, R. K. (2019). A novel method for parallel indexing of real time geospatial big data generated by IoT devices. *Future generation computer systems*, *97*, 433-452.

52. Liu, C., Feng, Y., Lin, D., Wu, L., & Guo, M. (2020). Iot based laundry services: an application of big data analytics, intelligent logistics management, and machine learning techniques. *International Journal of Production Research*, *58*(17), 5113-5131.

53. Liu, X., Tamminen, S., Su, X., Siirtola, P., Röning, J., Riekki, J., ... & Soininen, J. P. (2018, March). Enhancing veracity of IoT generated big data in decision making. In *2018 IEEE*

International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) (pp. 149-154). IEEE.

54. Lokshina, I. V., Durkin, B. J., & Lanting, C. J. (2017). Data analysis services related to the IoT and big data: strategic implications and business opportunities for third parties. *International Journal of Interdisciplinary Telecommunications and Networking (IJITN)*, *9*(2), 37-56.

55. López-Martínez, F., Núñez-Valdez, E. R., García-Díaz, V., & Bursac, Z. (2020). A case study for a big data and machine learning platform to improve medical decision support in population health management. *Algorithms*, *13*(4), 102.

56. Lu, J., Liu, A., Song, Y., & Zhang, G. (2020). Data-driven decision support under concept drift in streamed big data. *Complex & Intelligent Systems*, *6*(1), 157-163.

57. Maghsoodi, A. I., Riahi, D., Herrera-Viedma, E., & Zavadskas, E. K. (2020). An integrated parallel big data decision support tool using the W-CLUS-MCDA: A multi-scenario personnel assessment. *Knowledge-Based Systems*, *195*, 105749.

58. Maksimovic, M. (2017). The role of green internet of things (G-IoT) and big data in making cities smarter, safer and more sustainable. *International Journal of Computing and Digital Systems*, *6*(04), 175-184.

59. Massaro, A., Ricci, G., Selicato, S., Raminelli, S., & Galiano, A. (2020, June). Decisional Support System with Artificial Intelligence oriented on Health Prediction using a Wearable Device and Big Data. In *2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT* (pp. 718-723). IEEE.

60. Misra, N. N., Dixit, Y., Al-Mallahi, A., Bhullar, M. S., Upadhyay, R., & Martynenko, A. (2020). IoT, big data and artificial intelligence in agriculture and food industry. *IEEE Internet of Things Journal*.

61. Nair, L. R., Shetty, S. D., & Shetty, S. D. (2018). Applying spark based machine learning model on streaming big data for health status prediction. Computers & Electrical Engineering, 65, 393-399.

62. Oguntimilehin, A., & Ademola, E. O. (2014). A review of big data management, benefits and challenges. *A Review of Big Data Management, Benefits and Challenges*, *5*(6), 1-7.

63. Özemre, M., & Kabadurmus, O. (2020). A big data analytics based methodology for strategic decision making. Journal of Enterprise Information Management.

64. Polyakova, A., Loginov, M., Strelnikov, E., & Usova, N. (2019). Managerial decision support algorithm based on network analysis and big data. *International Journal of Civil Engineering and Technology*, *10*(2), 291-300.

65. Popchev, I., Orozova, D., & Stoyanov, S. (2019, November). IoT and Big Data Analytics in E-Learning. In *2019 Big Data, Knowledge and Control Systems Engineering (BdKCSE)* (pp. 1-5). IEEE.

66. Puthal, D., Ranjan, R., Nepal, S., & Chen, J. (2017). IoT and big data: An architecture with data flow and security issues. In *Cloud infrastructures, services, and IoT systems for smart cities* (pp. 243-252). Springer, Cham.

67. Qaffas, A. A., Hoque, R., & Almazmomi, N. (2021). The internet of things and big data analytics for chronic disease monitoring in Saudi Arabia. *Telemedicine and e-Health*, *27*(1), 74-81.

68. Rashid, M., Singh, H., Goyal, V., Parah, S. A., & Wani, A. R. (2021). Big data based hybrid machine learning model for improving performance of medical Internet of Things data in healthcare systems. In *Healthcare Paradigms in the Internet of Things Ecosystem* (pp. 47-62). Academic Press.

69. Rathore, M. M., Paul, A., Ahmad, A., & Jeon, G. (2017). IoT-based big data: from smart city towards next generation super city planning. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *13*(1), 28-47.

70. Rathore, M. M., Paul, A., Ahmad, A., Anisetti, M., & Jeon, G. (2017). Hadoop-based intelligent care system (HICS) analytical approach for big data in IoT. *ACM Transactions on Internet Technology (TOIT)*, *18*(1), 1-24.

71. Renjith, S., Sreekumar, A., & Jathavedan, M. (2021). SMaRT: A Framework for Social Media Based Recommender for Tourism. In *Second International Conference on Networks and Advances in Computational Technologies* (pp. 297-307). Springer, Cham

72. Sasubilli, S. M., Kumar, A., & Dutt, V. (2020, June). Improving Health Care by Help of Internet of Things and Bigdata Analytics and Cloud Computing. In *2020 International Conference on Advances in Computing and Communication Engineering (ICACCE)* (pp. 1-4). IEEE.

73. Sbai, I., & Krichen, S. (2020). A real-time Decision Support System for Big Data Analytic: A case of Dynamic Vehicle Routing Problems. *Procedia Computer Science*, *176*, 938-947.

74. Schrage, M. (2016). How the big data explosion has changed decision making. *Harvard Business Review*, *25*

75. Sestino, A., Prete, M. I., Piper, L., & Guido, G. (2020). Internet of Things and Big Data as enablers for business digitalization strategies. *Technovation*, 102173.

76. Sethi, P., & Sarangi, S. R. (2017). Internet of things: architectures, protocols, and applications. *Journal of Electrical and Computer Engineering*, *2017*.

77. Shah, S. A., Seker, D. Z., Rathore, M. M., Hameed, S., Yahia, S. B., & Draheim, D. (2019). Towards disaster resilient smart cities: Can internet of things and big data analytics be the game changers?. *IEEE Access*, *7*, 91885-91903.

78. Shamim, S., Zeng, J., Shariq, S. M., & Khan, Z. (2019). Role of big data management in enhancing big data decision-making capability and quality among Chinese firms: A dynamic capabilities view. *Information & Management*, *56*(6), 103135

79. Shukla, S., Kukade, V., & Mujawar, S. (2015). Big data: concept, handling and challenges: an overview. *International Journal of Computer Applications*, *114*(11), 6-9.

80. Si, C., & Shi, W. (2021, August). Establishment and Improvement of Financial Decision Support System Using Artificial Intelligence and Big Data. In *Journal of Physics: Conference Series* (Vol. 1992, No. 3, p. 032082). IOP Publishing.

81. Singh, P., Kaur, A., Batth, R. S., Kaur, S., & Gianini, G. (2021). Multi-disease big data analysis using beetle swarm optimization and an adaptive neuro-fuzzy inference system. *Neural Computing and Applications*, 1-12

82. Spiliopoulos, G., Zissis, D., & Chatzikokolakis, K. (2017, September). A big data driven approach to extracting global trade patterns. In *International Workshop on Mobility Analytics for Spatio-temporal and Social Data* (pp. 109-121). Springer, Cham

83. Srivastava, U., & Gopalkrishnan, S. (2015). Impact of big data analytics on banking sector: Learning for Indian banks. *Procedia Computer Science*, *50*, 643-652

84. Stein, N., Meller, J., & Flath, C. M. (2018). Big data on the shop-floor: sensor-based decision-support for manual processes. *Journal of Business Economics*, *88*(5), 593-616.

85. Stewart, R., & Velupillai, S. (2021). Applied natural language processing in mental health big data. *Neuropsychopharmacology*, *46*(1), 252

86. Sun, N., Morris, J. G., Xu, J., Zhu, X., & Xie, M. (2014). iCARE: A framework for big data-based banking customer analytics. *IBM Journal of Research and Development*, *58*(5/6), 4-1

87. Syafrudin, M., Alfian, G., Fitriyani, N. L., & Rhee, J. (2018). Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing. *Sensors*, *18*(9), 2946.

88. Syeda-Mahmood, T. (2018). Role of big data and machine learning in diagnostic decision support in radiology. *Journal of the American College of Radiology*, *15*(3), 569-576.

89. Taher, N. C., Mallat, I., Agoulmine, N., & El-Mawass, N. (2019, April). An IoT-Cloud based solution for real-time and batch processing of big data: Application in healthcare. In *2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART)* (pp. 1-8). IEEE.

90. Toapanta, S. M. T., Romero, H. X. H., Saltos, N. S. M., & Gallegos, L. E. M. (2019, July). Prototype of a security model applied to IoT with big data for tourist management in the cities of Ecuador. In *2019 Third World Conference on Smart Trends in Systems Security and Sustainablity (WorldS4)* (pp. 147-152). IEEE.

91. Towe, R., Dean, G., Edwards, L., Nundloll, V., Blair, G., Lamb, R., ... & Manson, S. (2020). Rethinking data-driven decision support in flood risk management for a big data age. *Journal of Flood Risk Management*, *13*(4), e12652.

92. Unal, D., Hammoudeh, M., Khan, M. A., Abuarqoub, A., Epiphaniou, G., & Hamila, R. (2021). Integration of federated machine learning and blockchain for the provision of secure big data analytics for Internet of Things. *Computers & Security*, *109*, 102393.

93. ur Rehman, M. H., Ahmed, E., Yaqoob, I., Hashem, I. A. T., Imran, M., & Ahmad, S. (2018). Big data analytics in industrial IoT using a concentric computing model. *IEEE Communications Magazine*, *56*(2), 37-43.

94. Verma, N., Malhotra, D., & Singh, J. (2020). Big data analytics for retail industry using MapReduce-Apriori framework. *Journal of Management Analytics*, *7*(3), 424-442

95. Vieira, A. A., Dias, L. M., Santos, M. Y., Pereira, G. A., & Oliveira, J. A. (2020). On the use of simulation as a Big Data semantic validator for supply chain management. *Simulation Modelling Practice and Theory*, *98*, 101985

96. Volk, M., Bosse, S., Bischoff, D., & Turowski, K. (2019, June). Decision-support for selecting big data reference architectures. In *International Conference on Business Information Systems* (pp. 3-17). Springer, Cham.

97. Voss, S., Sebastian, H. J., & Pahl, J. (2017). Introduction to intelligent decision support and big data for logistics and supply chain management minitrack. Proceedings of the 50th Hawaii International Conference on System Sciences

98. Wang, R. Y., Lin, P. A., Chu, J. Y., Tao, Y. H., & Ling, H. C. (2021). A decision support system for Taiwan's forest resource management using Remote Sensing Big Data. *Enterprise Information Systems*, 1-22.

99. Wang, T., Bhuiyan, M. Z. A., Wang, G., Rahman, M. A., Wu, J., & Cao, J. (2018). Big data reduction for a smart city's critical infrastructural health monitoring. IEEE Communications Magazine, 56(3), 128-133

100. Wang, W. Y. C., & Wang, Y. (2020). Analytics in the era of big data: the digital transformations and value creation in industrial marketing

101. Wang, Y., Kung, L., Wang, W. Y. C., & Cegielski, C. G. (2018). An integrated big data analytics-enabled transformation model: Application to health care. Information & Management, 55(1), 64-79

102. Watkins, D. (2021). Real-Time Big Data Analytics, Smart Industrial Value Creation, and Robotic Wireless Sensor Networks in Internet of Things-based Decision Support Systems. *Economics, Management, and Financial Markets*, *16*(1), 31-41.

103. Yang, Y., Zheng, X., Guo, W., Liu, X., & Chang, V. (2018). Privacy-preserving fusion of IoT and big data for e-health. *Future Generation Computer Systems*, *86*, 1437-1455.

104. Yasmin, M., Tatoglu, E., Kilic, H. S., Zaim, S., & Delen, D. (2020). Big data analytics capabilities and firm performance: An integrated MCDM approach. *Journal of Business Research*, *114*, 1-15

105. Yassine, A., Singh, S., Hossain, M. S., & Muhammad, G. (2019). IoT big data analytics for smart homes with fog and cloud computing. *Future Generation Computer Systems*, *91*, 563-573.

106. Zhan, Y., & Tan, K. H. (2020). An analytic infrastructure for harvesting big data to enhance supply chain performance. *European Journal of Operational Research*, *281*(3), 559-574

107. Zhao, J. D., & Ming, Z. H. A. O. (2018). A management decision support model of smart campus based on big data center. *DEStech Transactions on Computer Science and Engineering*, (icmsa).